Xác suất có điều kiện (Conditional Probability)

Thống kê máy tính và ứng dụng (ĐTTX)

Vũ Quốc Hoàng (vqhoang@fit.hcmus.edu.vn)
FIT - HCMUS

Ngày 6 tháng 3 năm 2022

Nội dung

1. Xác suất có điều kiện
2. Công thức nhân xác suất
3. Công thức xác suất toàn phần
4. Định lý Bayes
5. Các biến cố độc lập
6. Mô hình xác suất "lặp lại thí nghiệm độc lập"
7. Độc lập có điều kiện
8. Phương pháp xấp xỉ xác suất bằng mô phỏng
9. Ví dụ minh họa - Bài toán Monty Hall

Nội dung

1. Xác suất có điều kiện
2. Công thức nhân xác suất
3. Công thức xác suất toàn phần
4. Định lý Bayes
5. Các biến cố độc lập
6. Mô hình xác suất "lặp lại thí nghiệm độc lập"
7. Độc lập có điều kiện
8. Phương pháp xấp xỉ xác suất bằng mô phỏng
9. Ví dụ minh họa - Bài toán Monty Hall

Xác suất có điều kiện

Ở vòng chung kết AFF Cup 2021, xét các biến cố

$$\begin{cases} A : \text{Đội đương kim vô địch Việt Nam vô địch,} \\ B : \text{Đội mạnh Thái Lan vô địch,} \\ C : \text{Đội chủ nhà Singapore vô địch.} \end{cases}$$

- Trước vòng bảng: $P(A) > P(B) > P(C)$.
- Sau vòng bảng: $P(A)$ giảm, $P(C)$ tăng.
- Sau vòng bán kết: $P(A) = 0$, $P(B)$ tăng mạnh, $P(C) = 0$.
- Sau trận chung kết lượt đi: $P(B) \approx 1$.
- Sau trận chung kết lượt về: $P(B) = 1$.

Xác suất có điều kiện

Cần điều chỉnh, cập nhật xác suất (khả năng xảy ra) của các biến cố liên quan đến thí nghiệm T khi có thêm thông tin về T

- Thông tin về T được thể hiện bằng việc biết (các) biến cố nào đó đã xảy ra.

Xác suất của biến cố A khi biết biến cố B đã xảy ra được gọi là **xác suất có điều kiện** (conditional probability) của A khi biết B xảy ra, kí hiệu là $P(A|B)$ và được tính bằng định nghĩa

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ (với } P(B) > 0).$$

- $A \cap B$ là "A trong B",
- Chia cho $P(B)$ giúp chuẩn hóa xác suất.

Xác suất có điều kiện - Tính chất

Với biến cố B cho trước và $P(B) > 0$, $P(.|B)$ có thể hiểu là xác suất "tính trong không gian mẫu mới" B và là một độ đo xác suất hợp lệ

- $0 \leq P(A|B) \leq 1$,
- $P(\Omega|B) = 1$,
- $P(\varnothing|B) = 0$,
- $P(A^c|B) = 1 - P(A|B)$,
- $P(A \cup C|B) = P(A|B) + P(C|B) - P(A \cap C|B)$,
- $P(A|C, B) = P(A|C \cap B) = \frac{P(A \cap C|B)}{P(C|B)}$ (với $P(C|B) > 0$), ...

Hơn nữa, $P(E) = P(E|\Omega)$ với mọi biến cố $E \subset \Omega$.

Xác suất có điều kiện - Ví dụ 1

Tung một đồng xu đồng chất 3 lần, $\Omega = \{HHH, HHT, ..., TTT\}$

- Biến cố "lần 1 được ngửa": $B_1 = \{HHH, HHT, HTH, HTT\}$
- Biến cố "lần 2 được ngửa": $B_2 = \{HHH, HHT, THH, THT\}$
- Biến cố "được đúng 2 lần ngửa": $A = \{HHT, HTH, THH\}$
- Biến cố "được ít nhất 2 lần ngửa": $C = \{HHT, HTH, THH, HHH\}$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{8}, \qquad P(C) = \frac{|C|}{|\Omega|} = \frac{4}{8},$$

$$P(A|B_1) = \frac{P(A \cap B_1)}{P(B_1)} = \frac{|A \cap B_1|}{|B_1|} = \frac{|\{HHT, HTH\}|}{|B_1|} = \frac{2}{4} = P(A|B_2),$$

$$P(A|B_1, B_2) = P(A|B_1 \cap B_2) = \frac{P(A \cap B_1 \cap B_2)}{P(B_1 \cap B_2)} = \frac{|\{HHT\}|}{|\{HHH, HHT\}|} = \frac{1}{2},$$

$$P(C|B_1, B_2) = P(C|B_1 \cap B_2) = \frac{P(C \cap B_1 \cap B_2)}{P(B_1 \cap B_2)} = \frac{|\{HHT, HHH\}|}{|\{HHH, HHT\}|} = 1.$$

Xác suất có điều kiện - Ví dụ 2

Chọn ngẫu nhiên 1 bóng đèn

- A: "bóng đèn được chọn thỏa tiêu chí kiểu dáng",
- B: "bóng đèn được chọn thỏa tiêu chí độ sáng".

	Kiểu dáng		
Độ sáng	Đạt	Không đạt	**Tổng**
Đạt	117	3	120
Không đạt	8	2	10
Tổng	125	5	130

$$P(A) = \frac{125}{130}, \quad P(B) = \frac{120}{130}, \quad P(A \cap B) = \frac{117}{130},$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{117/130}{120/130} = \frac{117}{120} \text{ hoặc từ bảng ta có } P(A|B) = \frac{117}{120},$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{117/130}{125/130} = \frac{117}{125} \text{ hoặc từ bảng ta có } P(B|A) = \frac{117}{125}.$$

Xác suất có điều kiện - Ví dụ 3

Một hộp gồm 8 bi trắng và 2 bi đỏ. Lần lượt bốc 2 viên không hoàn lại. Tính xác suất "lần hai bốc được bi đỏ" biết "lần một bốc được bi trắng"?

- Cách giải thông thường: Không gian mẫu là các chỉnh hợp chọn 2 từ 10 bi. Gọi A, B lần lượt là các biến cố "lần hai bốc được bi đỏ", "lần một bốc được bi trắng". Ta có

$$P(B) = \frac{P_8^1 \times P_9^1}{P_{10}^2} = \frac{8 \times 9}{10 \times 9}, \quad P(A \cap B) = \frac{P_8^1 \times P_2^1}{P_{10}^2} = \frac{8 \times 2}{10 \times 9},$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{8 \times 2}{8 \times 9} = \frac{2}{9}.$$

- Cách giải vi diệu!: Khi lần một bốc được bi trắng thì trong hộp còn 7 bi trắng và 2 bi đỏ. Do đó xác suất để lần hai bốc được bi đỏ là

$$P(A|B) = \frac{2}{9}.$$

Nội dung

1. Xác suất có điều kiện
2. Công thức nhân xác suất
3. Công thức xác suất toàn phần
4. Định lý Bayes
5. Các biến cố độc lập
6. Mô hình xác suất "lặp lại thí nghiệm độc lập"
7. Độc lập có điều kiện
8. Phương pháp xấp xỉ xác suất bằng mô phỏng
9. Ví dụ minh họa - Bài toán Monty Hall

Công thức nhân xác suất

Công thức nhân xác suất (multiplication rule)

$$P(A \cap B) = P(B)P(A|B) \text{ (khi } P(B) > 0),$$

$$P(A \cap B) = P(A)P(B|A) \text{ (khi } P(A) > 0).$$

Trong nhiều trường hợp, xác suất có điều kiện $P(A|B)$ dễ tính hơn xác suất $P(A \cap B)$.

Công thức nhân tổng quát. Giả sử có n biến cố $A_1, ..., A_n$ với $P(A_1 \cap ... \cap A_n) > 0$, ta có

$$P(A_1 \cap ... \cap A_n) = P(A_1) \times P(A_2|A_1) \times P(A_3|A_1, A_2) \times ... \times P(A_n|A_1, A_2, ..., A_{n-1}).$$

Công thức nhân xác suất - Ví dụ

Một hộp gồm 8 bi trắng và 2 bi đỏ. Lần lượt bốc 3 viên không hoàn lại. Tính xác suất "lần một và lần hai bốc được bi trắng còn lần ba bốc được bi đỏ"?

Giải. Gọi A_i là biến cố "lần thứ i bốc được bi trắng" ($i = 1, 2, 3$). Khi đó A_i^c là biến cố "lần thứ i bốc được bi đỏ". Ta có

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3^c) &= P(A_1) \times P(A_2|A_1) \times P(A_3^c|A_1, A_2) \\ &= \frac{8}{10} \times \frac{7}{9} \times \frac{2}{8} \\ &= \frac{7}{45}. \end{aligned}$$

Nội dung

1. Xác suất có điều kiện
2. Công thức nhân xác suất
3. Công thức xác suất toàn phần
4. Định lý Bayes
5. Các biến cố độc lập
6. Mô hình xác suất "lặp lại thí nghiệm độc lập"
7. Độc lập có điều kiện
8. Phương pháp xấp xỉ xác suất bằng mô phỏng
9. Ví dụ minh họa - Bài toán Monty Hall

Công thức xác suất toàn phần

$B_1, B_2, ..., B_n$ được gọi là một **họ đầy đủ** các biến cố (hay một **phân hoạch**) của Ω nếu

1. $B_i \cap B_j = \varnothing, \forall i \neq j$,
2. $\Omega = B_1 \cup B_2 \cup ... \cup B_n$.

Công thức xác suất toàn phần (law of total probability). Giả sử có phân hoạch $B_1, B_2, ..., B_n$ với $P(B_i) > 0$ $(i = 1, ..., n)$, khi đó

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(B_i)P(A|B_i).$$

Đặc biệt

$$P(A) = P(B)P(A|B) + P(B^c)P(A|B^c).$$

Công thức xác suất toàn phần - Ví dụ

Một hộp gồm 8 bi trắng và 2 bi đỏ. Lần lượt bốc 2 viên không hoàn lại. Tính xác suất "lần hai bốc được bi đỏ"?

Giải. Gọi A_i là biến cố "lần thứ i bốc được bi trắng" ($i = 1, 2$). Khi đó A_i^c là biến cố "lần thứ i bốc được bi đỏ". Ta có

$$\begin{aligned} P(A_2^c) &= P(A_1)P(A_2^c|A_1) + P(A_1^c)P(A_2^c|A_1^c) \\ &= \frac{8}{10} \times \frac{2}{9} + \frac{2}{10} \times \frac{1}{9} \\ &= \frac{18}{90}. \end{aligned}$$

Nội dung

1. Xác suất có điều kiện
2. Công thức nhân xác suất
3. Công thức xác suất toàn phần
4. Định lý Bayes
5. Các biến cố độc lập
6. Mô hình xác suất "lặp lại thí nghiệm độc lập"
7. Độc lập có điều kiện
8. Phương pháp xấp xỉ xác suất bằng mô phỏng
9. Ví dụ minh họa - Bài toán Monty Hall

Định lý Bayes

Giả sử bạn đi xét nghiệm một bệnh nan y và được kết quả là **dương tính** (positive).

Biết rằng

- **Độ nhạy** (sensitive) của xét nghiệm là 90%: nghĩa là, trong 100 người bị bệnh thì khoảng 90 người cho kết quả xét nghiệm dương tính.
- **Độ đặc hiệu** (specificity) của xét nghiệm là 95%: nghĩa là, trong 100 người không bệnh thì khoảng 95 người cho kết quả xét nghiệm âm tính.
- **Độ phổ biến** (prevalence) của bệnh là 1/10000: nghĩa là, trong 10000 người thì có khoảng 1 người bị bệnh.

Vậy bạn có nên lo chuẩn bị hậu sự không?

- Trước khi xét nghiệm: xác suất bạn bị bệnh là 1/10000.
- Sau khi xét nghiệm: do kết quả dương tính, xác suất bạn bị bệnh sẽ tăng, nhưng là bao nhiêu?

Định lý Bayes

Định lý Bayes (Bayes' theorem, Bayes's rule). Giả sử có phân hoạch $B_1, B_2, ..., B_n$ với $P(B_i) > 0$ $(i = 1, ..., n)$ và biến cố A với $P(A) > 0$, khi đó, với mọi $i = 1, ..., n$ ta có

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)}$$

- $P(B_i)$: **xác suất tiên nghiệm** (prior probability) của B_i,
- $P(B_i|A)$: **xác suất hậu nghiệm** (posterior probability) của B_i khi biết A,
- $P(A|B_i)$: **xác suất hợp lý** (likelihood) của A theo B_i,
- Lưu ý, $P(A)$ không phụ thuộc vào B_i nên ta có

$$P(B_i|A) \propto P(B_i)P(A|B_i) \text{ (kí hiệu } \propto \text{ là "tỉ lệ với").}$$

Định lý Bayes - Ví dụ

Trong thí nghiệm xét nghiệm trên. Đặt các biến cố

- A: "bạn bị bệnh",
- B: "bạn xét nghiệm ra dương tính".

Ta có

- $P(A|B) = 0.9$,
- $P(A^c|B^c) = 0.95 \Rightarrow P(A|B^c) = 1 - P(A^c|B^c) = 1 - 0.95 = 0.05$,
- $P(B) = 1/10000 = 0.0001 \Rightarrow P(B^c) = 1 - P(B) = 1 - 0.0001 = 0.9999$.

Áp dụng định lý Bayses ta có

$$\begin{aligned} P(B|A) &= \frac{P(B)P(A|B)}{P(B)P(A|B) + P(B^c)P(A|B^c)} = \frac{0.0001 \times 0.9}{0.0001 \times 0.9 + 0.9999 \times 0.05} \\ &\approx 18/10000. \end{aligned}$$

Tính xác suất hậu nghiệm qua nhiều giai đoạn

Nếu thí nghiệm được tiến hành qua nhiều giai đoạn thì xác suất hậu nghiệm ở giai đoạn trước là xác suất tiên nghiệm của giai đoạn tiếp theo.

Ví dụ: giả sử, để chắc ăn, bạn xét nghiệm một lần nữa và vẫn ra dương tính

- Trước khi xét nghiệm: xác suất bạn bị bệnh là 1/10000,
- Sau khi xét nghiệm lần 1: do kết quả dương tính, xác suất bạn bị bệnh là 18/10000,
- Sau khi xét nghiệm lần 2: do kết quả vẫn dương tính, xác suất bạn bị bệnh sẽ tăng, nhưng là bao nhiêu?

Ta có $P(B) = 18/10000 = 0.0018 \Rightarrow P(B^c) = 1 - P(B) = 0.9982$.

Áp dụng định lý Bayses ta có

$$P(B|A) = \frac{P(B)P(A|B)}{P(B)P(A|B) + P(B^c)P(A|B^c)} = \frac{0.0018 \times 0.9}{0.0018 \times 0.9 + 0.9982 \times 0.05}$$
$$\approx 314/10000.$$

Nội dung

1. Xác suất có điều kiện
2. Công thức nhân xác suất
3. Công thức xác suất toàn phần
4. Định lý Bayes
5. Các biến cố độc lập
6. Mô hình xác suất "lặp lại thí nghiệm độc lập"
7. Độc lập có điều kiện
8. Phương pháp xấp xỉ xác suất bằng mô phỏng
9. Ví dụ minh họa - Bài toán Monty Hall

Các biến cố độc lập

Tung ngẫu nhiên một đồng xu đồng chất 2 lần, $\Omega = \{HH, HT, TH, TT\}$. Làm sao hình thức hóa trực quan "lần hai ngẫu nhiên so với lần một" hay "lần hai không phụ thuộc (về xác suất) lần một" hay "lần hai độc lập (về xác suất) lần một"?

- Đặt A_i là biến cố "tung được mặt ngửa ở lần $i = 1, 2$"

$$A_1 = \{HH, HT\}, \quad A_2 = \{HH, TH\}, \quad A_1 \cap A_2 = \{HH\}.$$

- Ta có $P(A_1) = \frac{|A_1|}{|\Omega|} = \frac{1}{2}, P(A_2) = \frac{|A_2|}{|\Omega|} = \frac{1}{2}, P(A_1 \cap A_2) = \frac{|A_1 \cap A_2|}{|\Omega|} = \frac{1}{4}$.
- Do đó $P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)} = \frac{1/4}{1/2} = \frac{1}{2} = P(A_2)$.
- Tương tự ta cũng có $P(A_2|A_1^c) = P(A_2)$ và $P(A_2^c|A_1) = P(A_2^c|A_1^c) = P(A_2^c)$.
- Vậy xác suất ra mặt nào ở lần hai cũng không thay đổi dù ra mặt nào ở lần một.

Các biến cố độc lập

Hai biến cố $\{A, B\}$ được gọi là **độc lập** (independent, statistically independent, stochastically independent) với nhau nếu

$$P(A \cap B) = P(A) \times P(B).$$

Một cách tương đương: $P(A|B) = P(A)$ $(P(B) > 0)$ hay $P(B|A) = P(B)$ $(P(A) > 0)$.

Mệnh đề. Nếu $\{A, B\}$ độc lập thì các cặp biến cố $\{A^c, B\}, \{A, B^c\}, \{A^c, B^c\}$ cũng độc lập.

Ở ví dụ tung đồng xu trên ta có $P(A_1 \cap A_2) = \frac{1}{4} = P(A_1) \times P(A_2)$ nên $\{A_1, A_2\}$ độc lập. Từ đó ta cũng có các cặp biến cố $\{A_1, A_2^c\}, \{A_1^c, A_2\}, \{A_1^c, A_2^c\}$ độc lập.

Các biến cố độc lập - Ví dụ

Tung một xúc xắc đồng chất, $\Omega = \{1, 2, 3, 4, 5, 6\}$.
Đặt biến cố

- "Gieo được mặt chẵn": $A = \{2, 4, 6\}$,
- "Gieo được mặt không quá 4": $B = \{1, 2, 3, 4\}$,
- "Gieo được mặt không quá 3": $C = \{1, 2, 3\}$.

Ta có

$$P(A \cap B) = \frac{2}{6} = P(A) \times P(B) = \frac{3}{6} \times \frac{4}{6},$$

$$P(A \cap C) = \frac{1}{6} \neq P(A) \times P(C) = \frac{3}{6} \times \frac{3}{6}.$$

Vậy $\{A, B\}$ độc lập nhưng $\{A, C\}$ không độc lập. Giải thích trực quan?

Các biến cố độc lập

Ba biến cố $\{A, B, C\}$ được gọi là độc lập (với nhau) nếu từng đôi $\{A, B\}, \{A, C\}, \{B, C\}$ độc lập và

$$P(A \cap B \cap C) = P(A) \times P(B) \times P(C).$$

Ví dụ: tung ngẫu nhiên một đồng xu đồng chất 2 lần, $\Omega = \{HH, HT, TH, TT\}$.
Đặt biến cố

- "Tung được mặt ngửa ở lần một": $A = \{HH, HT\}$,
- "Tung được mặt ngửa ở lần hai": $B = \{HH, TH\}$,
- "Tung được mặt giống nhau ở hai lần": $C = \{HH, TT\}$.

Hãy cho thấy 3 biến cố $\{A, B, C\}$ độc lập từng đôi nhưng cả 3 không độc lập với nhau.
Giải thích trực quan?

Nội dung

1. Xác suất có điều kiện
2. Công thức nhân xác suất
3. Công thức xác suất toàn phần
4. Định lý Bayes
5. Các biến cố độc lập
6. **Mô hình xác suất "lặp lại thí nghiệm độc lập"**
7. Độc lập có điều kiện
8. Phương pháp xấp xỉ xác suất bằng mô phỏng
9. Ví dụ minh họa - Bài toán Monty Hall

Mô hình xác suất "lặp lại thí nghiệm độc lập"

Họ các biến cố $\{A_1, A_2, ...\}$ được gọi là độc lập nếu với mọi tập con khác rỗng và hữu hạn $\{B_1, B_2, ..., B_k\}$ của họ ta có

$$P\left(\bigcap_{i=1}^{k} B_i\right) = \prod_{i=1}^{k} P(B_i).$$

Lưu ý, định nghĩa tính độc lập trên có thể được dùng theo hai chiều

- Chiều ngược: dùng để cho thấy (hay kiểm tra, chứng minh) tính độc lập,
- Chiều xuôi: dùng giả thuyết về tính độc lập để tính toán xác suất đơn giản.

Chẳng hạn, chiều xuôi, được dùng trong **mô hình xác suất "lặp lại thí nghiệm độc lập"**: thực hiện lặp lại thí nghiệm T nhiều lần một cách độc lập, gọi A_i là biến cố "liên quan đến lần thực hiện thứ i" thì

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i).$$

Ví dụ

Cho một đồng xu không đồng chất với xác suất ra ngửa là 0.7. Tung đồng xu (một cách độc lập) đến khi ra mặt sấp thì dừng. Tính xác suất phải tung đồng xu 10 lần?

- Đặt A_i là biến cố "tung được mặt ngửa ở lần i" ($i = 1, 2, ...$).
- Ta có $P(A_i) = 0.7$ và $P(A_i^c) = 1 - 0.7 = 0.3$ với mọi $i = 1, 2, ...$
- Theo giả thuyết độc lập ta có họ $\{A_1, A_2, ...\}$ độc lập.
- Từ đó ta có xác suất phải tung đồng xu 10 lần là

$$P\left(\left(\bigcap_{i=1}^{9} A_i\right) \cap A_{10}^c\right) = \left(\prod_{i=1}^{9} P(A_i)\right) P(A_{10}^c) = 0.7^9 \times 0.3 = 0.0121.$$

- Lưu ý, không gian mẫu là vô hạn $\Omega = \{T, HT, HHT, HHHT, ...\}$.

Nội dung

1. Xác suất có điều kiện
2. Công thức nhân xác suất
3. Công thức xác suất toàn phần
4. Định lý Bayes
5. Các biến cố độc lập
6. Mô hình xác suất "lặp lại thí nghiệm độc lập"
7. **Độc lập có điều kiện**
8. Phương pháp xấp xỉ xác suất bằng mô phỏng
9. Ví dụ minh họa - Bài toán Monty Hall

Độc lập có điều kiện

Cho biến cố C với $P(C) > 0$, họ các biến cố $\{A_1, A_2, ...\}$ được gọi là **độc lập có điều kiện** (conditionally independent) khi biết C nếu với mọi tập con khác rỗng và hữu hạn $\{B_1, B_2, ..., B_k\}$ của họ ta có

$$P\left(\bigcap_{i=1}^{k} B_i | C\right) = \prod_{i=1}^{k} P(B_i|C).$$

Hai biến cố $\{A, B\}$ được gọi là độc lập có điều kiện khi biết C nếu

$$P(A \cap B|C) = P(A|C) \times P(B|C).$$

Một cách tương đương:

$$P(A|B, C) = P(A|C)\ (P(B|C) > 0) \text{ hay } P(B|A, C) = P(B|C)\ (P(A|C) > 0).$$

Độc lập có điều kiện - Ví dụ

Tung một xúc xắc đồng chất, $\Omega = \{1, 2, 3, 4, 5, 6\}$.
Đặt biến cố

- "Gieo được mặt chẵn": $A = \{2, 4, 6\}$,
- "Gieo được mặt không quá 4": $B = \{1, 2, 3, 4\}$,
- "Gieo được mặt không quá 3": $C = \{1, 2, 3\}$.
- "Gieo được mặt không quá 2": $D = \{1, 2\}$.

Từ ví dụ trước, ta đã biết $\{A, B\}$ độc lập nhưng $\{A, C\}$ không độc lập. Ta có

$$P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{1/6}{2/6} = \frac{1}{2},$$

$$P(A|C, D) = \frac{P(A \cap C \cap D)}{P(C \cap D)} = \frac{1/6}{2/6} = \frac{1}{2}.$$

Vậy $\{A, C\}$ độc lập có điều kiện khi biết D. Giải thích trực quan?

Nội dung

1. Xác suất có điều kiện
2. Công thức nhân xác suất
3. Công thức xác suất toàn phần
4. Định lý Bayes
5. Các biến cố độc lập
6. Mô hình xác suất "lặp lại thí nghiệm độc lập"
7. Độc lập có điều kiện
8. **Phương pháp xấp xỉ xác suất bằng mô phỏng**
9. Ví dụ minh họa - Bài toán Monty Hall

Phương pháp xấp xỉ xác suất bằng mô phỏng

Để xấp xỉ xác suất có điều kiện của một biến cố B khi biết biến cố A đã xảy ra trong thí nghiệm T, ta có thể dùng phương pháp thống kê như sau

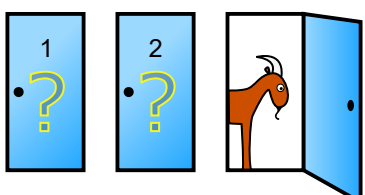
- Thực hiện lặp lại N lần (độc lập) thí nghiệm T, đếm số lần biến cố A xảy ra, m, và trong các lần A xảy ra thì đếm số lần B cũng xảy ra, p. Khi đó $f(B|A) = \frac{p}{m}$ được gọi là tần suất của B trên A.
- Khi N đủ lớn, ta có $f(B|A) \approx P(B|A)$.
- Việc thực hiện lặp lại nhiều lần thí nghiệm T có thể được **mô phỏng** (simulation) trên máy tính.

Nội dung

1. Xác suất có điều kiện
2. Công thức nhân xác suất
3. Công thức xác suất toàn phần
4. Định lý Bayes
5. Các biến cố độc lập
6. Mô hình xác suất "lặp lại thí nghiệm độc lập"
7. Độc lập có điều kiện
8. Phương pháp xấp xỉ xác suất bằng mô phỏng
9. Ví dụ minh họa - Bài toán Monty Hall

Bài toán Monty Hall

Monty Hall problem. Có 3 cửa #1, #2, #3 chứa 1 chiếc xe và 2 con dê. Người chơi chọn 1 cửa (chẳng hạn #1). Người dẫn chương trình biết cửa nào có gì. Người dẫn chọn và mở cửa có dê trong 2 cửa còn lại (chẳng hạn #3). Người dẫn hỏi người chơi có muốn đổi lựa chọn không (vẫn giữ #1 hay chọn #2). Người chơi nên giữ hay đổi (để khả năng được xe cao hơn)?



(https://en.wikipedia.org/wiki/Monty_Hall_problem)

Bài toán Monty Hall - Tính toán chính xác

Do "tính đối xứng" nên ta có thể giả sử kịch bản như mô tả (người chơi chọn cửa #1, người dẫn mở cửa #3).

Đặt A_i là biến cố "xe được đặt ở cửa #i" ($1 \leq i \leq 3$) và B_j là biến cố "người dẫn mở cửa #j" ($1 \leq j \leq 3$). Từ bài toán và kịch bản đã cho (người chơi chọn cửa #1), ta có

- $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$.
- $P(B_3|A_1) = \frac{1}{2}$ (người chơi đã chọn cửa #1, xe cũng được đặt ở cửa #1 nên người dẫn có thể mở 1 trong 2 cửa #2, #3).
- $P(B_3|A_2) = 1$ (người chơi đã chọn cửa #1, xe được đặt ở cửa #2 nên người dẫn chỉ có thể mở cửa #3).
- $P(B_3|A_3) = 0$ (xe được đặt ở cửa #3 nên người dẫn không được mở cửa #3).

Bài toán Monty Hall - Tính toán chính xác (tt)

Dùng công thức Bayes, ta có xác suất người chơi được xe khi không đổi cửa là

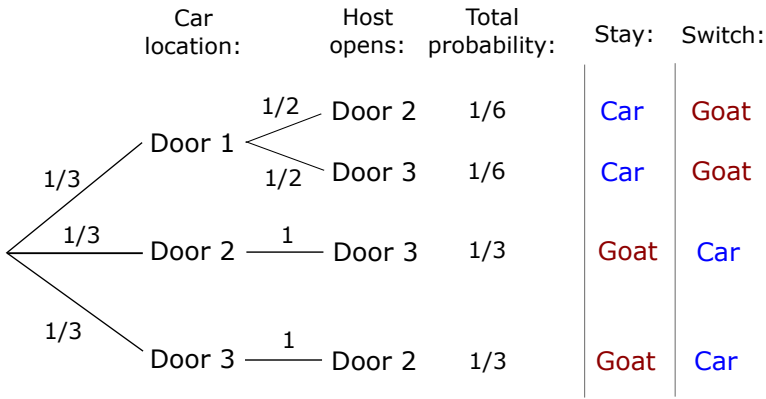
$$\begin{aligned} P(A_1|B_3) &= \frac{P(A_1)P(B_3|A_1)}{P(A_1)P(B_3|A_1) + P(A_2)P(B_3|A_2) + P(A_3)P(B_3|A_3)} \\ &= \frac{\frac{1}{3}\frac{1}{2}}{\frac{1}{3}\frac{1}{2} + \frac{1}{3}1 + \frac{1}{3}0} = \frac{1}{3} \end{aligned}$$

và xác suất người chơi được xe khi đổi cửa là

$$\begin{aligned} P(A_2|B_3) &= \frac{P(A_2)P(B_3|A_2)}{P(A_1)P(B_3|A_1) + P(A_2)P(B_3|A_2) + P(A_3)P(B_3|A_3)} \\ &= \frac{\frac{1}{3}1}{\frac{1}{3}\frac{1}{2} + \frac{1}{3}1 + \frac{1}{3}0} = \frac{2}{3}. \end{aligned}$$

Vậy người chơi nên chọn đổi cửa (xác suất được xe cao gấp đôi so với không đổi cửa).

Bài toán Monty Hall - Sơ đồ cây (tree diagram)



(https://en.wikipedia.org/wiki/Monty_Hall_problem#Conditional_probability_by_direct_calculation)

Bài toán Monty Hall - Mô phỏng

```
import random

def Monty_Hall(doors={"#1", "#2", "#3"}):
    car_door = random.choice(list(doors))
    choice_door = random.choice(list(doors))
    open_door = random.choice(list(doors - {choice_door,
       car_door}))
    op_door = random.choice(list(doors - {choice_door,
       open_door}))

    return car_door == choice_door, car_door == op_door
```

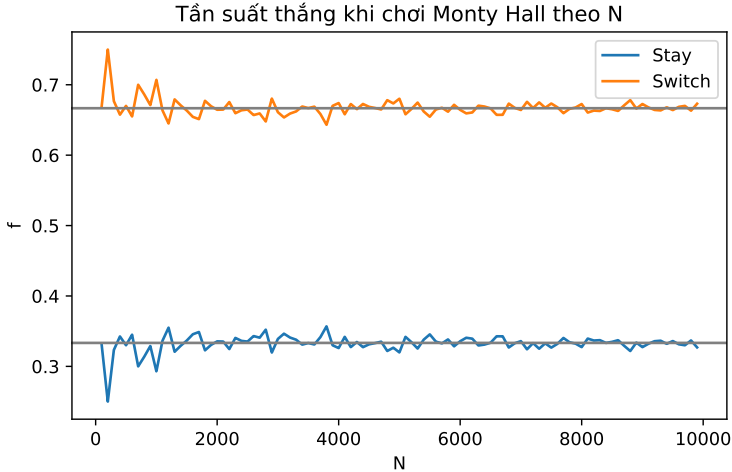
Bài toán Monty Hall - Mô phỏng (tt)

```
N = 10000
results = [Monty_Hall() for _ in range(N)]

stay_freq = sum([v for v, _ in results])/N
switch_freq = sum([v for _, v in results])/N

print(stay_freq, switch_freq)
#0.3317 0.6683
```

Bài toán Monty Hall - Mô phỏng (tt)



Tài liệu tham khảo

Chapter 2. Morris H. DeGroot, Mark J. Schervish. *Probability and Statistics*. Addison-Wesley, 2012.

Chapter 1, 2. H. Pishro-Nik. *"Introduction to probability, statistics, and random processes"*, available at `https://www.probabilitycourse.com`. Kappa Research LLC, 2014.